Response to Anonymous Referee #2

Received and published: 28 October 2020

The study proposed a novel method to capture the best matching unit in a self- organizing maps algorithm. The authors proved that this method is better suited for clustering meteorological data than a more widely used Euclidean distance method. This is an interesting and well written manuscript.

There are some minor housekeeping corrections / suggestions that the author can find in the attached pdf. Other than that the manuscript is acceptable to be published in my point of view.

Please also note the supplement to this comment: https://gmd.copernicus.org/preprints/gmd-2020-278/gmd-2020-278-RC2- supplement.pdf

We appreciate the reviewer for the constructive and helpful comments for improving the quality of this paper. We have carefully addressed all these comments, conducting additional tests and analysis suggested by the reviewer. The manuscript has also been revised accordingly.

Before going into further details, we would like to keep the reviewer notified of **one miscoding we found in line 73** of python script "*silhouette_anal.py*" (to calculate the silhouette scores), **but we confirmed that this miscoding does not largely affect the conclusions**. The original code was "*abis = np.argsort(ab)[::-1]*". The correct code must be "*abis = np.argsort(ab)*". This miscoding affected term b in the silhouette-score formula: $s = (b - a)/(\max{(a, b)})$, where b is the mean distance between a sample and all other points in the *next nearest cluster*. However, with original code, b became the mean distance between a sample and all other points in the *farthest cluster*. Consequently, silhouette scores had been incorrectly overestimated. We have corrected the python script and re-calculated all the silhouette scores. The correction has changed the score's absolute values, but it likely does not affect this study's conclusion. With the newly calculated silhouette score, S-SOM still exhibits the consistently higher performance over ED-SOM.

Comments:

Line 78 There are more than one method to find the BMU. Usually, Euclidean or Karl Pearson distances are used, but distance can be defined by any measure that might be appropriate to the particular problem. Here the authors introduce a new one and CHOSED one other popular method to compared to. It would be better to state that clearer since there are other methods that can deal with structural data, for example using correlation coefficient.

We have examined the performance of SOM using the Pearson correlation coefficient (i.e., "structure" in S-SIM when c_3=0), hereafter called COR-SOM. As the reviewer speculated, the results show that COR-SOM performed better than ED-SOM (in terms of both the silhouette score and topographic error), particularly when cluster number is small (see Fig. R1, 2 below, or Fig. 6, 7 in the main text). Indeed, COR-SOM has comparable performance to S-SOM in experiments MAM and SON. However, COR-SOM is lower scored than S-SOM in DJF and partly in JJA.

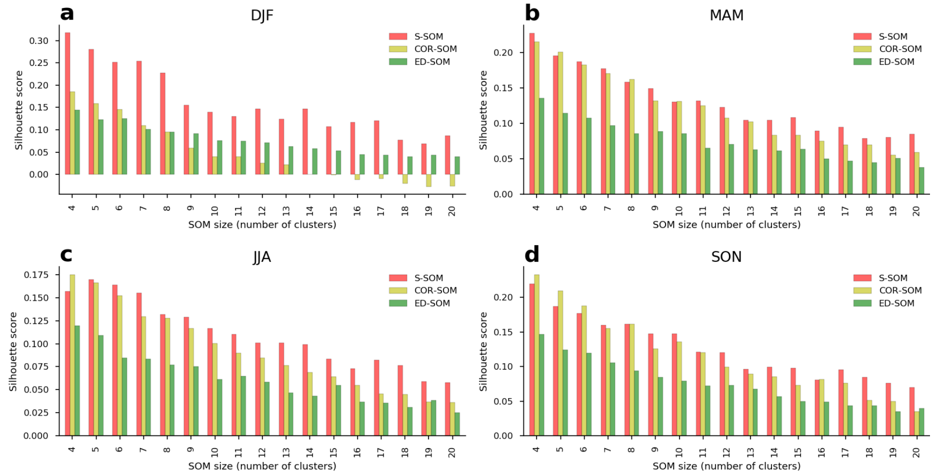

Fig. R1. Silhouette scores of S-, COR- and ED-SOM. a) the winter experiment DJF; b) spring MAM; c) summer JJA, and d) autumn SON. In each plot, the x-axis denotes different SOM size configurations, and the y-axis represents the silhouette score that ranges from -1 to 1, with 1 means the perfect cluster assignment, 0 means a sample is located at the edge, a negative value means wrong cluster assignment.

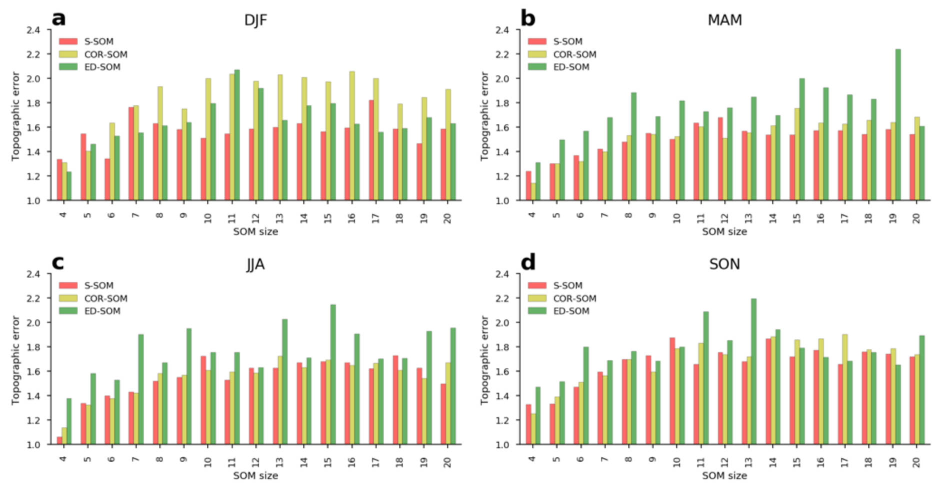

Fig. R2. Topographic errors of S-, COR- and ED-SOM. a) the results for winter DJF; b) for spring MAM; c) for summer JJA, and d) for autumn SON. In each plot, the x-axis denotes different SOM size configurations, and the y-axis denotes topographic errors, with the minimum value of 1 indicating “no error” or best topographical preservation.